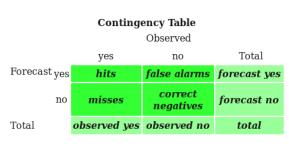
Verification of predictions of CME arrival time at L1

Ed Pope, Met Office, UK

Verifying predictions of CME arrival time at L1

- Compared MOSWOC archived forecasts & CME Scoreboard average of methods with the Scoreboard observed time
- Data: April-December 2014
- Method:
 - Compare MOSWOC arrival time prediction with observed arrival time on Scoreboard.
 - Produce a MOSWOC contingency table (hit, miss, false alarm, correct rejections).
 - Do same for Scoreboard average.
 - Calculate scores & confidence intervals for both approaches.
- Confidence interval: a range of values likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data.
- Confidence levels (e.g. 95%): if the same population is sampled on numerous occasions & interval estimates are made on each occasion, the resulting intervals would bracket the true population parameter in ~95% of cases.

ME: 2016-04-10T11:00:00-CME-001												
Actual Shock Arrival Time: 2016-04-14T06-50Z												
Observed Geomagnetic Storm Parameters: Max Kp: 5.0 CME Note: CME associated with large filament eruption situated close to N18E29 starting around 10UTC.												
Predicted Shock Arrival Time	<u>Difference</u> (hrs)	Confidence (%)	Submitted On	<u>Lead Time</u> (hrs)	Predicted Geomagnetic Storm Parameter(s)	<u>Method</u>	Submitted By					
2016-04-14T00:00Z (-7.0h, +7.0h)	-6.83		2016-04-11T00:54Z	77.93		WSA-ENLIL + Cone (GSFC SWRC)	Yaireska Collado (GSFC)	Detail				
2016-04-13T14:00Z	-16.83		2016-04-11T05:07Z	73.72	Max Kp Range: 5.0	WSA-ENLIL + Cone (NOAA/SWPC)	Leila Mays (GSFC)	<u>Detail</u>				
016-04-13T18:00Z (-12.0h, +6.0h)	-12.83	30.0	2016-04-11T05:45Z	73.08	Max Kp Range: 4.0 - 6.0	WSA-ENLIL + Cone (Met Office)	Met Office (Met Office)	Detail				
2016-04-14T12.00Z (-12.0h, +12.0h)	5.17		2016-04-11T12:30Z	66.33	****	Other (SIDC)	Leila Mays (GSFC)	Detail				
2016-04-13T04:51Z	-25.98	100.0	2016-04-12T20:30Z	34.33		SPM2	Xinhua Zhao (NSSC CAS)	<u>Detail</u>				
2016-04-13T12:44Z	-18.10		2016-04-12T20:33Z	34.28		SPM	Xinhua Zhao (NSSC CAS)	Detail				
2016-04-13T18:15Z	-12.58	65.0			Max Kp Range: 4.0 - 5.5	Average of all Methods	Auto Generated (CCMC)	<u>Detail</u>				
						·						



_				Score-			
Score	140014/00	E0/ 61	050/ 61	board	E0/ OI	050/ 61	
Hits	MOSWOC 33	5% CL	95% CL	averg 27	5% CL		A measure of
Misses	9			0			Number of times a yes forecast was a yes occurrence. Number of times a no forecast was a yes occurence.
False alarms	6			12			Number of times a restorecast was a yes occurence.
Correct rejections	7			9			Number of times a yes forecast was a no occurrence.
Correct rejections	,						Number of times a no forecast was a no occurrence.
							Discrimination
							1=perfect
							S=perfect.
Hit rate	0.79	0.68	0.88	1	1		Ranges do not overlap.
THETALE	0.75	0.08	0.00		1		Discrimination
False alarm rate	0.46	0.23	0.7	0.57	0.4		S>M, however ranges overlap.
raise alaim rate	0.46	0.23	0.7	0.57	0.4		Reliability
							!
False alarm ratio	0.15	0.07	0.25	0.01	0.40		S is significantly higher than M.
	0.15		0.25				Ranges just overlap.
Probability of detection	0.6			0.56			
Probability of false detection	0.12	0.05	0.18	0.25	0.17		
							Accuracy
		0.54	0.00		0.55		Fraction of hits & correct rejections.
Proportion correct	0.73		0.82	0.75	0.65		Comparable for both.
Base rate	0.76		0.86				
Forecast rate	0.71	0.6	0.8	0.8	0.73		
							Accuracy
- 1 .		0.50	0.70		0.57		0=no skill, 1=perfect
Threat score	0.69	0.58	0.79	0.69	0.57	0.81	Comparable for both.
							Bias
							1=perfect
							M<1 so under-forecasting. S>1 so over-forecasting.
Bias score	0.93	0.79	1.09	1.44	1.24	1.76	Ranges overlap.
							Skill
							Accounts for hits occurring by chance in the threat score.
							0=no skill, 1=perfect.
Equitable threat score	0.18	0.04	0.34	0.3	0.16		Two approaches are comparable & ranges overlap.
							Skill
							Fractional improvement over just chance.
_							M slightly lower than S & ranges ovelap.
Heidke score	0.3	0.07	0.51	0.46	0.27	0.64	Suggests some skill in both forecasting approaches.
							Skill
							Similar to Heidke.
Peirce score	0.32	0.08	0.57	0.43	0.25	0.6	The two approaches are comparable & ranges overlap.

Results

Summary

- Only a short period of data analysed rerun with more data, preferably several years
 - may help to reduce confidence intervals
 - as indication of whether skill has changed over time (improved through experience/ got worse through losing STEREO?)
- Difficult to strongly distinguish differences between MOSWOC & Scoreboard average.
- Suggestion that NASA are over-predicting (high hit-rate & high false alarm rate).
- Ambiguity of 'hit' e.g. when CMEs in quick succession .
- Would be interesting to do cost-benefit analysis, since false alarms are potentially expensive for users.
- http://www.cawcr.gov.au/projects/verification/